

# LIPID MAPS online tools for lipid research

Eoin Fahy<sup>1</sup>, Manish Sud<sup>1</sup>, Dawn Cotter<sup>1</sup> and Shankar Subramaniam<sup>1,2,\*</sup>

<sup>1</sup>LIPID MAPS Bioinformatics Core, San Diego Supercomputer Center, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92037, USA and <sup>2</sup>Departments of Bioengineering, Chemistry and Biochemistry, University of California, San Diego, La Jolla, CA 92093, USA

Received January 31, 2007; Revised March 27, 2007; Accepted April 17, 2007

## ABSTRACT

The LIPID MAPS consortium has developed a number of online tools for performing tasks such as drawing lipid structures and predicting possible structures from mass spectrometry (MS) data. A simple online interface has been developed to enable an end-user to rapidly generate a variety of lipid chemical structures, along with corresponding systematic names and ontological information. The structure-drawing tools are available for six categories of lipids: (i) fatty acyls, (ii) glycerolipids, (iii) glycerophospholipids, (iv) cardiolipins, (v) sphingolipids and (vi) sterols. Within each category, the structure-drawing tools support the specification of various parameters such as chain lengths at a specific *sn* position, head groups, double bond positions and stereochemistry to generate a specific lipid structure. The structure-drawing tools have also been integrated with a second set of online tools which predict possible lipid structures from precursor-ion and product-ion MS experimental data. The MS prediction tools are available for three categories of lipids: (i) mono/di/triacylglycerols, (ii) glycerophospholipids and (iii) cardiolipins. The LIPID MAPS online tools are publicly available at [www.lipidmaps.org/tools/](http://www.lipidmaps.org/tools/).

## INTRODUCTION

The structures of large and complex lipids are difficult to represent in drawings, which leads to the use of many custom formats that often generate more confusion than clarity among members of the lipid research community. For example, usage of the Simplified Molecular Line Entry Specification (SMILES) (1) ([www.daylight.com/smiles/index.html](http://www.daylight.com/smiles/index.html)) format to represent lipid structures, while being very compact and accurate in terms of bond connectivity, valence and chirality, causes problems when the structure is rendered. This is due to the fact that the SMILES format does not include 2D coordinates and

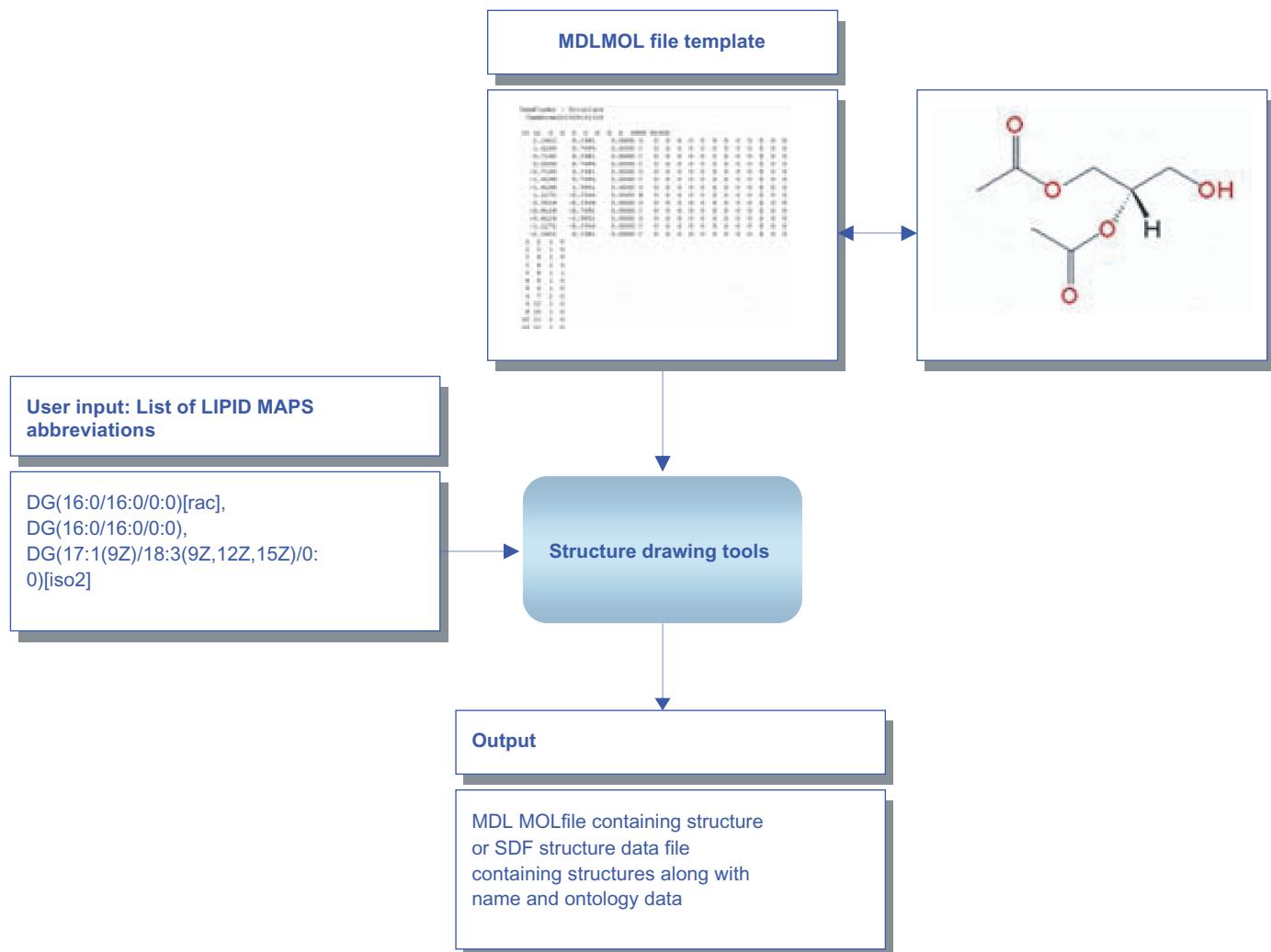
hence the orientation of the structure as drawn is quite arbitrary, making visual recognition and comparison of related structures difficult. Members of the lipid community currently draw structures based on their own individual preferences. A given lipid structure may appear quite differently in different lipid databases (2, 3). In summary, consistent structure-drawing tools for lipids are currently not available.

The structure-drawing step is typically a most time-consuming process in creating molecular databases of lipids. However, many classes of lipids lend themselves to automated structure-drawing paradigms, due to their consistent 2D layout. The LIPID MAPS consortium has developed and deployed a suite of structure-drawing tools that greatly increase the efficiency of data entry into lipid structure databases and permit ‘on-demand’ structure generation in conjunction with a variety of MS prediction tools. We have chosen a consistent format for representing lipid structures (4) where, in the simplest case of the fatty acid derivatives, the acid group (or equivalent) is drawn on the right and the hydrophobic hydrocarbon chain is on the left. Similarly for glycerolipids, glycerophospholipids and sphingolipids, the radyl hydrocarbon chains are drawn to the left and the headgroups are depicted on the right. This approach enables a more consistent, error-free approach to drawing lipid structures and has been used extensively in populating the LIPID MAPS structure database (LMSD), which currently contains over 10 000 molecules (5).

We have adopted an approach where ‘core’ structures such as diacetyl glycerol (glycerolipids) and formic acid (fatty acyls) are represented as text-based MDL molfiles (described under section MDL CTfile Formats at [www.mdli.com](http://www.mdli.com)), and these molfiles are then manipulated to generate a variety of structures in MDL molfile and Structure Data Format (SDF) files containing that core (Figure 1). This manipulation is carried out by command-line or online programs written in the Perl programming language.

The structural similarities of many lipid categories also make it feasible to predict structures from MS precursor ion and/or product ion data by creating a database composed of masses of all possible likely combinations of

\*To whom correspondence should be addressed. Tel: +1 858 822 0986; Fax: +1 858 822 3752; Email: shankar@sdsc.edu



**Figure 1.** Schematic demonstrating the principle of using molfile templates and a list of lipid abbreviations as input for structure-drawing tools.

acyl side chains for a given lipid core. One can then use matching algorithms to display possible candidates for given precursor ion/product ion  $m/z$  values and then generate corresponding structures.

## DESCRIPTION AND IMPLEMENTATION

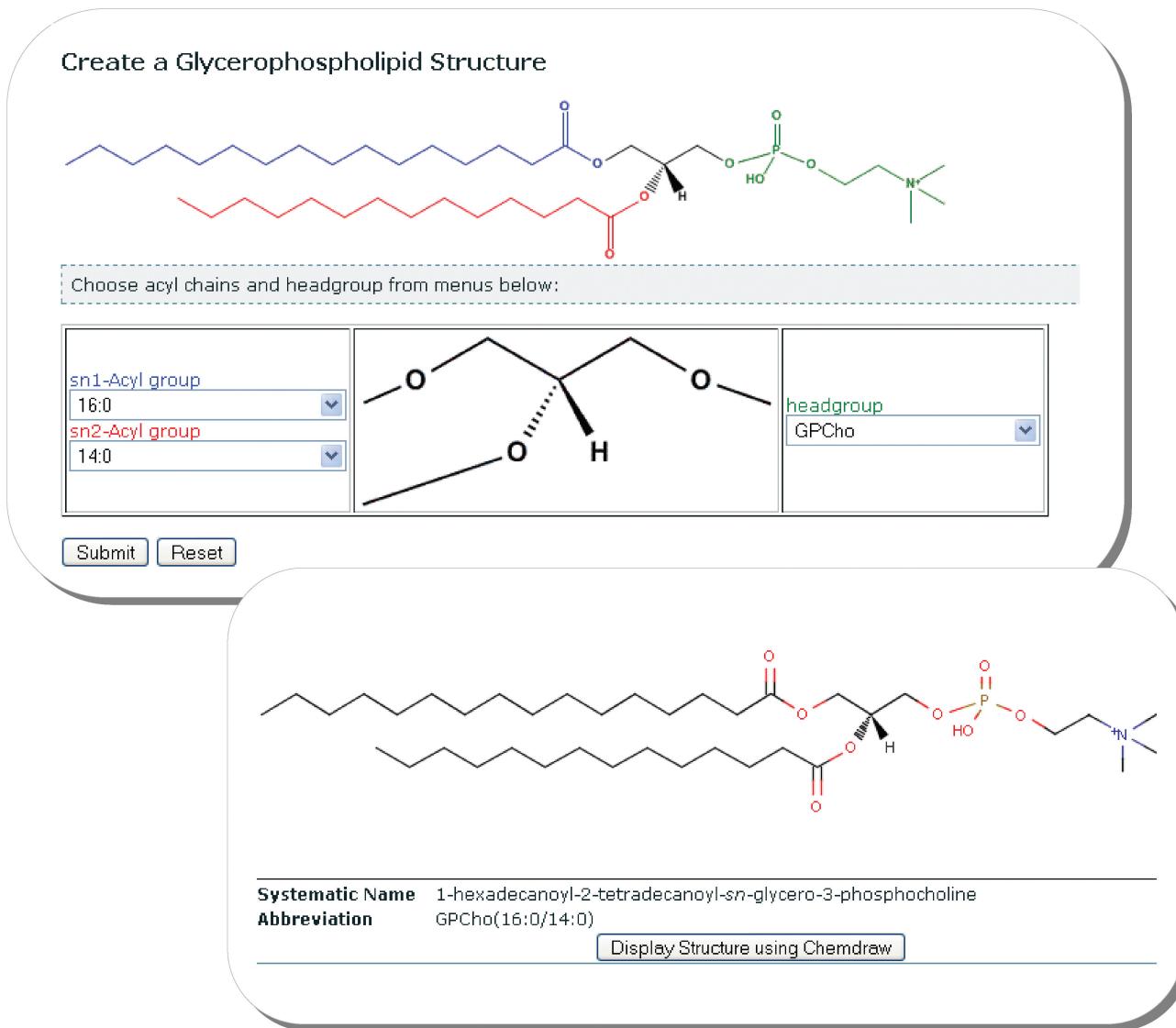
### Structure-drawing tools

The LIPID MAPS website ([www.lipidmaps.org/tools/index.html](http://www.lipidmaps.org/tools/index.html)) currently contains a suite of six structure-drawing tools for the following lipid categories: fatty acyls, glycerolipids, glycerophospholipids, cardiolipins, sphingolipids and sterols. The online layout (Figure 2) consists of a 'core' structure and pull-down menus arranged in locations appropriate for that structure. For example, in the case of the glycerophospholipid-drawing tool, a central glycerol core is surrounded by pull-down menus allowing the end-user to choose from a list of headgroups and  $sn1$  and  $sn2$  acyl side chains. The list of acyl chains represents the more common species found in mammalian cells, and could easily be modified to include

additional chains. The selected lipid structure is then generated via a server-side Perl script. The structure is rendered in the web browser as a Java-based MarvinView applet ([www.chemaxon.com/marvin/](http://www.chemaxon.com/marvin/)). Additionally, the structure may be viewed online with the Chemdraw ActiveX/Plugin ([www.cambridgesoft.com/software/ChemDraw/](http://www.cambridgesoft.com/software/ChemDraw/)) by users who have this component installed on their system. Current versions of the fatty-acyl-drawing tools are now capable of drawing chiral centers and ring structures. Molecules with correct stereochemistry are drawn by implementing the following method: (1) usage of the PerlMol ([www.perlmol.org/](http://www.perlmol.org/)) module to define atoms, bonds and neighbors; (2) a recursive algorithm which applies Cahn–Ingold–Prelog (CIP) (6, 7) rules to a chiral center and (3) a scoring system to estimate substituent priority to assign chirality.

### LIPID MAPS abbreviation

Concurrently, a generalized lipid abbreviation format has been developed which enables structures, systematic names and ontologies to be generated automatically



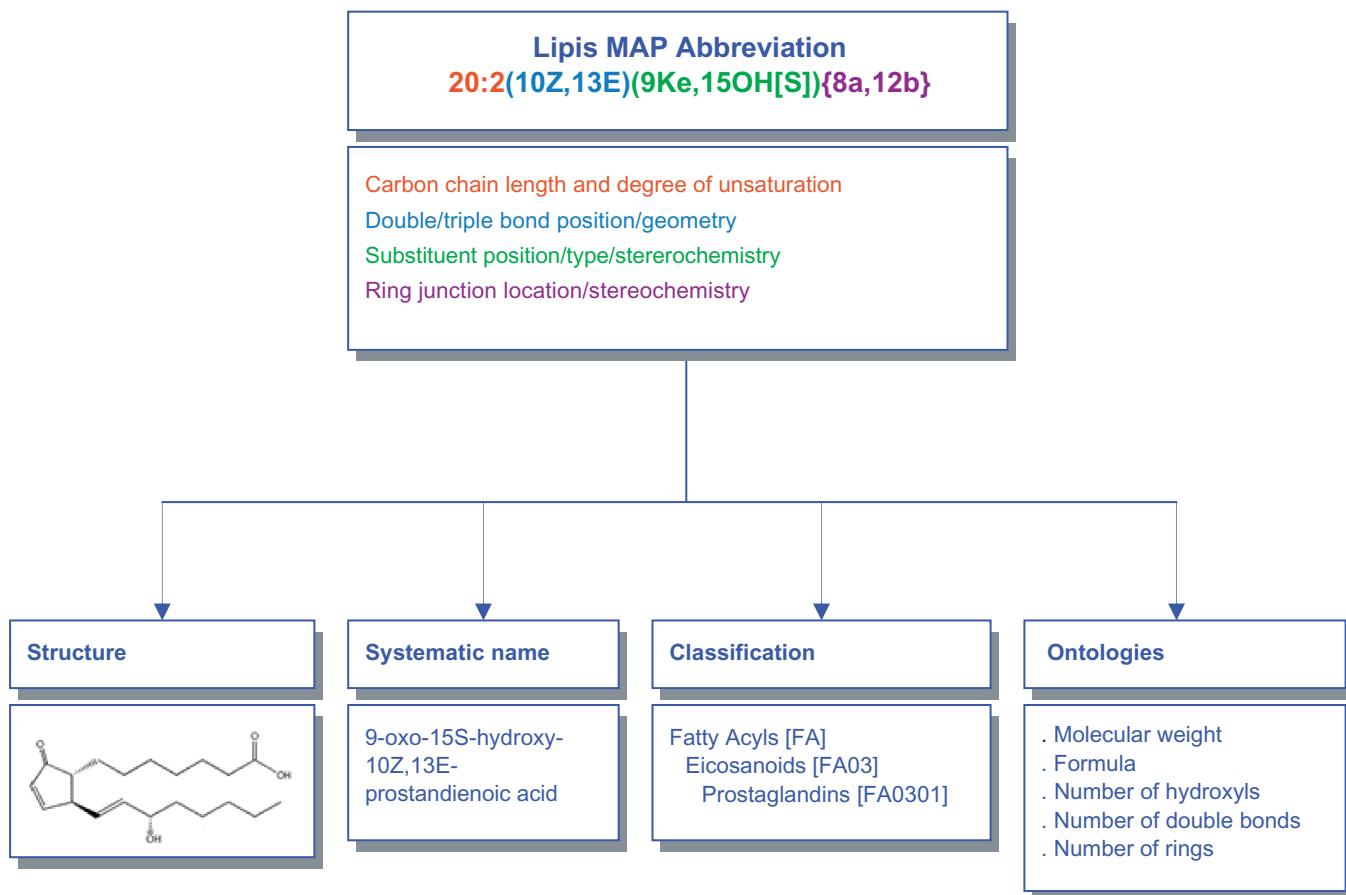
**Figure 2.** Online structure-drawing tool for glycerophospholipids.

from a single source format (Figure 3). The LIPID MAPS abbreviation format for lipids may consist of up to four different parts: (i) carbon chain length along with any degree of unsaturation; (ii) position and geometry of double and triple bonds; (iii) position, type and stereochemistry of substituents and (iv) position of carbocyclic ring junction and stereochemistry. The first part of the abbreviation format describing the carbon chain is mandatory; the other parts are optional. For example, the LIPID MAPS abbreviation, 20:2 (10Z, 13E) (9Ke, 15OH[S]) {8a,12b}, for a fatty acyl structure consists of four parts: (i) 20:2—20-carbon chain length with two double bonds; (ii) (10Z, 13E)—two double bonds at carbon positions 10 and 13 with Z (cis) and E (trans) double bond geometry; (iii) (9Ke, 15OH[S])—a keto group at carbon 9 and a hydroxyl group at carbon 15 with S stereochemistry and (iv) {8a,12b}—a ring junction at carbon positions 8 and 12 with alpha and beta stereochemistry at positions 8 and 12 respectively.

Using this approach, a text file containing a list of lipid abbreviations may be submitted in batch mode to a drawing application which then generates structures (as molfiles or SD files), systematic names and ontological information such as formula, molecular weight, number of rings, number of double/triple bonds, hydroxyl, amino, keto groups, etc.). In this way, thousands of lipid structures have been generated in a consistent fashion and deposited in the LIPID MAPS structure database with considerable savings in time. Furthermore, the associated ontological information has been databased and used in various online search interfaces where end-users may search for structures by presence (or number) of a functional group or other features.

#### MS prediction tools

Certain classes of lipids such as acylglycerols and phospholipids composed of an invariant core (glycerol and



**Figure 3.** Flowchart showing structure/name/ontology generation from an abbreviation of a lipid. This example demonstrates the conversion of a text abbreviation for a prostaglandin into a dataset containing structure (MDL Molfile), systematic name, classification and various molecular attributes such as formula, molecular weight, number of functional groups, double bonds and rings.

headgroups) and one or more acyl/alkyl substituents are good candidates for MS computational analysis. These molecules tend to fragment in a predictable fashion in collision-induced experiments leading to loss of acyl side chains, neutral loss of fatty acids and loss of water and other diagnostic ions (8) depending on the nature of the headgroup. It is possible to create a virtual database of permutations of the more common side chains for glycerolipids and glycerophospholipids and calculate ‘high-probability’ product ion candidates in order to compare the experimental data with predicted spectra. The LIPID MAPS group has developed a suite of search tools that allows a user to enter an  $m/z$  value of interest and view a list of matching structure candidates, along with a list of calculated neutral-loss ions and other ‘high-probability’ product ions. The MS prediction tools are currently available for three different categories of lipids: glycerolipids, cardiolipins and glycerophospholipids. In each case, all possible structures corresponding to a list of likely headgroups and acyl, alkyl-ether and vinyl-ether chains have been expanded and enumerated by computational methods to generate a table containing the nominal and exact mass for each discrete structure as well as

additional ontological information such as formula, abbreviation and numbers of chain carbons and double bonds. This tabular data is then uploaded into category-specific database tables, making it amenable for online tools.

MS prediction may also be extended by allowing the user to input a list of experimentally observed product ions, in addition to the precursor ion. An online search tool for glycerophospholipids has been implemented by computing product ion masses for commonly observed fragments corresponding to acyl chain ions, neutral loss of acyl chains, loss of water, headgroup-specific fragmentations and combinations of the above. Observations of common fragment ions were obtained from the LIPID MAPS online library of lipid standards ([www.lipidmaps.org/data/standards/](http://www.lipidmaps.org/data/standards/)), which includes tandem mass spectral data generated by the LIPID MAPS core facilities for over 90 glycerophospholipids and glycerolipids. The query interface accepts an  $m/z$  value for the precursor ion and a list of product ions which are then matched against the database entries to generate a list of likely structures (based on precursor ion match) and a ‘fragment score’ (based on the number of product ion matches).

## USER INTERFACES

### Structure-drawing tools

A simple online interface has been developed to enable an end-user to rapidly generate a variety of lipid chemical structures, along with corresponding systematic names and ontological information. The user interface is implemented using a combination of Perl and PHP scripts.

The fatty acyl structure-drawing tools support the specification of any valid fatty acyl LIPID MAPS abbreviation representing acyl chain length along with the specification of double bonds, triple bonds and a variety of substituents on the acyl chain. Alternatively, the user can also choose desired substituents for up to three positions on the acyl chain from the pull-down menus. Currently, the following substituents are supported: hydroxyl –OH, amino –NH<sub>2</sub>, thio - SH, methyl –Me, ethyl –Et, propyl –Pr, methoxy –OMe, acetoxy –OAc, keto/oxo –Ke, epoxy –Ep, methylene –My, hydroperoxy –OOH, bromo –Br, chloro –Cl, fluoro –F, cyano –CN, carboxy –COOH, aldehyde –CHO, and methyl ester – COOMe.

The glycerolipid structure-drawing tools provide the capability to generate a variety of monoradyl, diradyl and triradyl glycerols by specification of acyl chains at *sn*<sub>1</sub>, *sn*<sub>2</sub> and *sn*<sub>3</sub> positions. After the user chooses the acyl chain abbreviation for different positions from the pull-down lists, the corresponding lipid structure is generated. Over 50 different acyl chain abbreviations are available for *sn*<sub>1</sub>, *sn*<sub>2</sub> and *sn*<sub>3</sub> positions. Acyl chain abbreviation of 0:0 at the *sn*<sub>3</sub> and/or *sn*<sub>2</sub> positions are used to generate structures for diradyl and monoradyl glycerols.

The cardiolipin structure-drawing tools also provide a pull-down list of acyl chain abbreviations to choose for positions *sn*<sub>1(1')</sub>, *sn*<sub>2(1')</sub>, *sn*<sub>1(3')</sub>, and *sn*<sub>2(3')</sub>. Except for the absence of acyl chain abbreviation of 0:0, this list is similar to the list of acyl chain abbreviations for glycerolipids. After the user chooses desired acyl chain abbreviations for all the four positions, the corresponding cardiolipin structure is generated.

The glycerophospholipid structure-drawing tools, in addition to pull-down list of acyl chain abbreviations for *sn*<sub>1</sub> and *sn*<sub>2</sub> position also provide a pull-down list for various headgroups. Except for the absence of acyl chain abbreviation of 0:0 at *sn*<sub>1</sub> position, this list is similar to the list of acyl chain abbreviations for glycerolipids. The following headgroup abbreviations are supported: GPCho (glycerophosphocholines), GPSer (glycerophosphoserines), GPEtn (glycerophosphoethanolamines), GPA (glycerophosphates), GPGro (glycerophosphoglycerols) and GPIns (glycerophosphoinositols).

The sphingolipid structure-drawing tools support the generation of a wide variety of sphingolipid structures by providing pull-down lists for sphingoid base, N-acyl group and headgroups. Supported sphingoid bases are: sphingosine, sphinganine, C16 sphingosine, C16 sphinganine, C17 sphingosine, C17 sphinganine, C19 sphingosine and C19 sphinganine. N-acyl group pull-down list contains the following acyl chain abbreviations: 0:0, 10:0, 12:0, 14:0, 16:0, 18:0, 20:0, 22:0, 24:0, 24:1(5Z), 26:0, and 26:1(17Z).

The supported headgroups are: sphingomyelin, ceramide, ceramide-1-phosphate, and glucosyl ceramide.

The sterol-drawing tools support the generation of structures derived from cholestan sterol core. In addition to double bond position specification, the user can choose to substitute atoms in the cholestan core by C, N, O and H along with the stereochemistry specification of alpha or beta for the substituted atom. Pull-down lists for position, stereochemistry and atom specification are provided for up to four simultaneous substitutions.

### MS prediction tools

The MS prediction tools for glycerolipids, cardiolipins and glycerophospholipids accept an *m/z* value from the user for the precursor ion and have a menu to allow selection of the ion mode ([M+H]<sup>+</sup>, [M+NH4]<sup>+</sup>, [M-H]<sup>-</sup>, etc). In addition, a mass tolerance range and a headgroup (in the case of glycerophospholipids) may be specified to limit the number of matches. The list of matches may also be filtered by specifying a particular chain, using the LIPID MAPS abbreviation format. On completion of a search, the output format (Figure 4) contains a list of structures that (a) satisfy the input criteria and (b) whose side chains belong to the list (48 in the case of glycerolipids and glycerophospholipids, 10 in the case of cardiolipins) of radyl chains used to populate the database. The predicted masses of the fragment ions are computed at run-time by the online application. All entries in the result set are hyperlinked to the structure-drawing application (under the ‘Abbreviation’ column), enabling ‘on-demand’ visualization of the molecular structures. Isotopic distribution profiles for each structure may also be viewed under the ‘Formula’ column.

## SUMMARY AND FUTURE WORK

The LIPID MAPS bioinformatics group has developed a publicly available suite of structure-drawing tools for several major categories of lipids, with a view to improving the speed and consistency of drawing these structures. The drawing tools may be accessed via an online interface, or as standalone Perl scripts (for glycerolipids/glycerophospholipids). The standalone scripts may be used for creation of structures (as SDF files) in batch mode. In the future, the scope of both online and standalone drawing tools will be extended to accommodate additional lipid categories such as sterols and prenol lipids.

The structural similarities within certain lipid categories such as glycerolipids and glycerophospholipids has been exploited by creating online tools for predicting structures from MS precursor ion and/or product ion data by generating ‘virtual databases’ of possible lipid permutations. These search tools have been integrated with the aforementioned drawing tools to permit ‘on-demand’ structure generation. Since the LIPID MAPS consortium is heavily involved in collecting MS data from a wide variety of lipid species, these prediction tools will be improved and customized in future versions by validation with fragmentation data from experimental MS/MS spectra.

## Mass Spectrometry

### Show Possible Glycerophospholipid Structures

Enter a mass (m/z) value and corresponding ion type:

<input checked="" type="radio"/> Nominal Mass	<input type="radio"/> Exact Mass	671
Mass Tolerance (+/- amu) +/- 1.0 amu		
Radyl group abbreviation (eg. 12:0) contains		
Headgroup Any		
Ion M-H		
Sort by Molecular Formula		
Records per page 50		

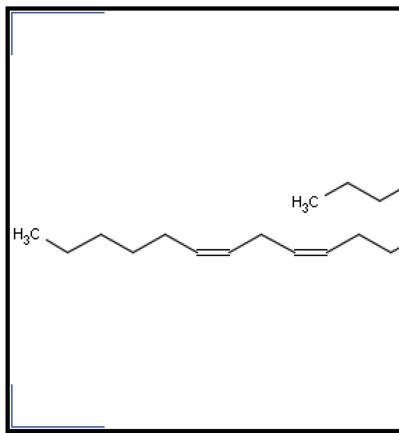
### Possible Glycerophospholipid Structures

Exact Mass: 671 Mass Tolerance: +/- 1.0 amu

C=Number of Carbons; DB=Number of double bonds; sn1('1),sn2...=MS/MS Product Ions (neutral loss)

Mass	C	DB	Abbrev.	M-sn1-H	M-sn1-H2O-H	M-sn2-H	M-sn2-H2O-H	sn1 acid(-)	sn2 acid(-)
671.4652	34	2	<a href="#">12:0/22:2(13Z,16Z)</a>	489.2981	471.2875	353.1729	335.1623	199.1698	335.295
671.4652	34	2	<a href="#">14:0/20:2(11Z,14Z)</a>	461.2668	443.2562	381.2042	363.1936	227.2011	307.2637
671.4652	34	2	<a href="#">16:0/18:2(9Z,12Z)</a>	471.2875	453.2769	381.2042	363.1936	227.2011	307.2637

### Possible Glycerophospholipid Structure



Systematic Name 1-dodecanoyl-2-(13Z,16Z)-sn-glycerophosphatidylcholine

Abbreviation 12:0/22:2(13Z,16Z)

Exact Mass 671.4652

Formula C<sub>37</sub>H<sub>69</sub>O<sub>8</sub>P

[View Structure using ChemDraw](#) (?)

### Isotopic Distribution for C<sub>37</sub>H<sub>69</sub>O<sub>8</sub>P

Total	m/z	% of Max	Correction Factor
M:	672.4730	100.0000	1.5471
M+1:	673.4764	42.5418	3.6366
M+2:	674.4797	10.2843	15.0431
M+3:	675.4831	1.7649	-
M+4:	676.4864	0.1161	-

### Isotopic Distribution for C<sub>37</sub>H<sub>69</sub>O<sub>8</sub>P

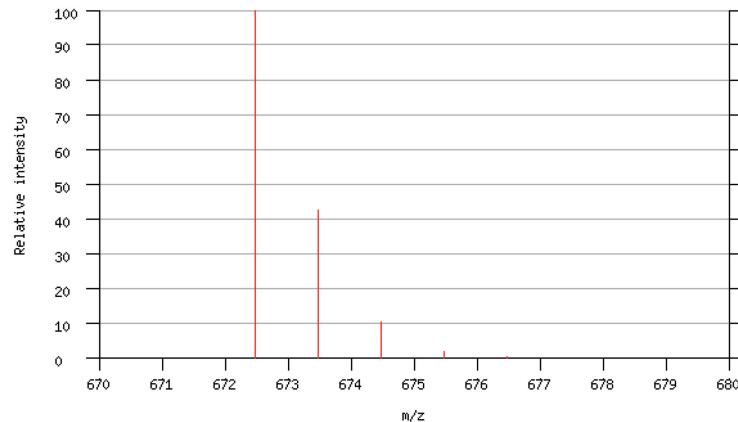


Figure 4. Online MS prediction tools for glycerophospholipids.

## ACKNOWLEDGEMENTS

This work was supported by National Institutes of Health (NIH) and National Institute of General Medical Sciences (NIGMS) Glue Grant NIH/NIGMS Grant 1 U54 GM69338. Funding to pay the Open Access publication charge was also provided by Glue Grant NIH/NIGMS Grant 1 U54 GM69338. The authors would like to thank Dr Robert Murphy, University of Colorado Health Sciences Ctr. for helpful suggestions and comments.

*Conflict of interest statement.* None declared.

## REFERENCES

1. Weininger,D. (1988) SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. *J. Chem. Inf. Comput. Sci.*, **28**(1), 31–36.
2. Watanabe,K., Yasugi,E. and Ohshima,M. (2000) How to search the glycolipid data in “Lipidbank for web”, the newly-developed lipid database in Japan. *Trends Glycosci. Glycotechnol.*, **12**, 175–184.
3. Caffrey,M. and Hogan,J. (1992) LIPIDAT: a database of lipid phase transition temperatures and enthalpy changes. DMPC data subset analysis. *Chem. Phys. Lipids*, **61**, 1–109.
4. Fahy,E., Subramaniam,S., Brown,H.A., Glass,C.K., Merrill,A.H.Jr., Murphy,R.C., Raetz,C.R., Russell, D.W. Seyama,Y. *et al.* (2005) A comprehensive classification system for lipids. *J. Lipid Res.*, **46**, 839–862.
5. Sud,M., Fahy,E., Cotter,D., Brown,H.A., Dennis,E.A., Glass,C.K., Merrill,A.H.Jr., Murphy,R.C., Raetz, *et al.* (2007) LMSD: LIPID MAPS structure database. *Nucleic Acids Res.*, **35**, D527–532 Database issue.
6. Cahn,R.S., Ingold,C. and Prelog,V. (1966) Specification of molecular chirality. *Angew. Chem. Int. Ed.*, **5**(4), 385–415.
7. Prelog,V. and Helmchen,G. (1982) Basic principles of the CIP-system and proposals for a revision. *Angew. Chem. Int. Ed.*, **21**, 567–583.
8. Murphy,R. C., Fiedler,J. and Hevko,J. (2001) Analysis of non-volatile lipids by mass spectrometry. *Chem. Rev.*, **101**, 479–526.